Classifying the Occupation Concepts in The Modern Arabic Lexicon

-Technical Report

Heba Qadah, Loay Mustafa and Nour Shayeb

Computer Science Department
Birzeit University

Abstract

There are a lot of occupations in this world, some of them still used, others disappeared and maybe there are also new occupations we still don't know about. So, it will be very useful if there is a source that collects these occupations in addition to classifying them. Therefore, in this project we made five contributions. Firstly, we collect the largest amount of terms in Arabic form the Lexicon of Modern Arabic in addition to their definitions, and then we categorize them as occupations or not. Secondly, we linked these terms with their corresponding classification from the Standard Occupational Classification (SOC)- "a system that classifies workers and jobs into occupational categories for the purpose of collecting, calculating, analyzing, or disseminating data"[1]. Thirdly, we Arabized the classifications as well by providing equivalent and relevant Arabic terms. Finally, we transformed the Arabic terms and their classifications into owl language so it can be used as extension to the Modern Arabic Lexicon and to be imported into the protégé.

I. Introduction

Ontology, originally a philosophical concept, it has gained more and more attention among researchers in the field of computer science recently, and it becomes more and more important as it is applied to different fields such as knowledge engineering, database design, information modeling, information query and so on[2], Ontology is the cornerstone for the exchange of information and data, as it contains an accurate definition of the semantic and conceptual meaning of the data to be exchanged, Ontology is not only given more attention in the field of computer science, but also in social and linguistic.

With the rapid development of the internet and data industry, personal and occupation queries are an urgent need [3][4]. Either for statistical research, setting salary scales [1] and enabling acknowledgement of different and the role each performs within the enterprise.

Hence the need to find Arabic ontology that is interested in professions and categorize them so that Arabic speakers can benefit from and easily access data from the ontology found in other languages.

Our work aims to Arabize an ontology of professions and link it with one of the Arabic dictionaries and upload it on Arab Ontology [11] developed at Birzeit University, in order to facilitate the access of the largest number of people. Arabizing a professions ontology is not all about text translation from the original language to the local language, but correct incorporation of meanings and synonyms in the local language and the mother language of the ontology through the use of dictionaries in the local language and linking them with the main ontology.

II. Related work

Occupational and Employment cluster (APA) ontology: This ontology includes many professions and classifies its classes into four main classes (Employee occupational job characteristics, Occupational groups, Organizations organizational behavior, Personnel management Professional Personnel Issues). [5]

The Role Ontology (RoleO): provides role classification and supports computer-assisted reasoning. RoleO development follows the OBO Foundry principles. RoleO has represented all the occupations defined in the Standard Occupational Classification (SOC) System from the USA Bureau of Labor Statistics.[6]

Standard Occupational Classification (SOC): is the United States government system for classifying occupations. It is used by the U.S. federal government agencies for collecting occupational data, enabling comparison of occupations across data sets. It is designed to cover all occupations in which work is performed for pay or profit, reflecting the current occupational structure in the United States. In the 2018 SOC, there are 98 minor groups. Each minor group is broken into broad occupations, of which there are 459 classes. There are, at the highest level of specification, 867 detailed occupations. Detailed occupations with similar job duties, and in some cases, similar skills, education, and/or training, are grouped together in the SOC. Each worker is classified into only one of the 867 detailed occupations based on the tasks he or she performs.[1]

III. Arabizing the Occupation Ontology Branch

First step in our work is to find good occupation ontologies that have the large number of professions that exist in the Arab world, to extract the appropriate data that helps in our work.

After a great effort in the research and discussion of a large number of ontologies, we use two ontologies: the APA Occupational and Employment cluster, and the Role Ontology.

One of the challenges we faced was the absence of parent classifications and many careers do not exist in this ontology. Thus, we used the Standard Occupational Classification in the United States, which contains 98 minor groups and 867 detailed occupations based on the tasks he or she performs.

Then we transformed the data from the above-mentioned ontologies and classified

and verified them according to the Standard Occupational Classification as shown in figure1.

lemma	occupational Group
مَنْ يحكم النَّاس ويتولَّى شئون إدارة	حاكم	Politicians
لقَب يُطلق على أبناء الملوك والأمر	أمير	Politicians
صفة مشبَّهة تدلّ على الثبوت من أ	أمير	Politicians
حاكم عام:- مَن يحكم منطقة كبيرة .	حاكم	Politicians
حاكم أعلى:- (السياسة) مَن يمارس	حاكم	Politicians
وظيفة الرئيس رئاسة الجمهوريّة :	رئاسة	Politicians
رئيس السُّلطة:- (السياسة) مصطلح	رئيس	Politicians
صفة مشبَّهة تدلّ على الثبوت من ر	رئيس	Politicians
مالك وسيّد "حوَ هُوَ كُلَّ عَلى مَوْلأَه	مَوْلَى	Politicians
لقَب رئاسيّ في بعض بلدان الخليج	شَيْخ	Politicians
رئيس القرية، خاصّةً في مصر، وذ	عُمْدَة	Politicians
زعيم، رئيس دولة عظمى "الأقطاب	قُطْب	Politicians
صيغة المؤنَّث لفاعل ناب1/ ناب إل	نائبة	Management_Personnel
مَنْ يُعتمد عليه في تدبير أمرٍ، شخ	وكيل	Management_Personnel
نائب "وكيل، النائب العام ـ فلانة وك	وكيل	Management_Personnel

Figure1: extracted data and the first classification.

IV. Arabic Lexical Concepts

The second step in our work is to find an Arabic language lexicon that contains a wide range of professions and jobs prevailing in the Arab world, and contains clear definitions.

We found that Lexicon of Modern Arabic (digitized by [8][9]) meets the conditions that we set previously, we extracted all the words in this lexicon and their semantic definitions (about 12600 dictionary entries). Then, we manually categorized those entries in the lexicon that denote occupations, which were approximately 600 entries (with there defnetions).

V. Translation Link between the lexicon and ontologies

We translate the classification, which was extracted in section 2 to Arabic language by using the synonyms. For this task we used the lexicons [8],[9],[10] in the Arabic ontology portal, developed at Birzeit university to find accurate translations. We then linked between the occupations we extract from the lexicon and our classification by setting the occupations and their synonyms and definition. After that, we upload all our work to Arabic Ontology [11] for Birzeit university[1], See figure 3. Figure 2 shows our Arabized ontology using Protégé.

Figure2: data after entering the protégé.

Figure3: data uploaded to Arabic ontology

VI. Challenges and problems

We encountered many problems that we can summarize as follows:

There are many standards and systems for classifying occupations, but there is a limited number of occupation ontologies.

The absence of an Arab ontology specialized in occupation.

The presence of a large number of occupations that are called by compound words that is not present in Arabic lexicon.

Occupational classifications in English are difficult to translate to Arabic and vice versa.

VII. CONCLUSIONS

We use two ontologies (APA Occupational and Employment cluster and The Role Ontology) and Standard Occupational Classification for the United States government to build new Arabized and enriched classification for occupations. We linked between this classification and the Arabic Lexicon to build Arabic ontology about the occupations.

VIII. REFERENCES

[1] Executive Office of The President Office of Management and Budget, Standard Occupational Classification Manual 2018, pp. 1-4.

[2] N. Guarino, "Formal Ontology and Information Systems," Proceedings of the 1st International Conference on Formal Ontology in Information Systems, IOS Press, 1998, pp.3-15.

[3] Abiteboul S., Agrawal R., Bernstein P., Carey M., Ceri S., Croft B., et al (2005). Proc. of the Joint Conference on Information Sciences, "The Lowell Database Research Self Assessment", Communications of the ACM, Vol.48, pp. 111-118.

[4] Yang B., Glen J. (2006). "Retroactive Answering of Search Queries", Proc. 15th International World Wide Web Conference. pp. 173-183 Edinburgh, Scotland.

[5] APA Occupational and Employment cluster, may 19,2020 from BioPortal. http://bioportal.bioontology.org/ontologies/APAOCUEMPLOY

[6] Role Ontology, may 19,2020 from BioPortal http://bioportal.bioontology.org/ontologies/ROLEO

[8] Jarrar, Mustafa & Amayreh, Hamzeh. (2019). An Arabic-Multilingual Database with a Lexicographic Search Engine. 234-246. 10.1007/978-3-030-23281-8_19.

[9] Mustafa Jarrar, Hamzeh Amayreh, John P. McCrae: Representing Arabic Lexicons in Lemon - a Preliminary Study. The 2nd Conference on Language, Data and Knowledge (LDK 2019). Pages(29--33),

Volume 2402, CEURS. Leipzig, Germany. 2019

[10] Diana Alhafi, Anton Deik, Mustafa Jarrar: Usability Evaluation of Lexicographic e-Services. The 16th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA). Pages(1--7), IEEE. Abu Dhabi, UAE. 2019

[11] Mustafa Jarrar: Building a Formal Arabic Ontology (Invited Paper). Proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks. ALECSO, Arab League. Tunis. July 2011